DOCUMENT RESUME

ED 081 782 TM 003 142

AUTHOR	Humphry, Betty
TITLE	Development and Analysis of a Taped and Written Test for Guidance Counselors: A Pilot Study.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-TDR-73-1
PUB DATE	Apr 73
NOTE	42p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Counselor Evaluation; Graduate Students; Guidance Personnel; Higher Education; *Multiple Choice Tests; Questionnaires; Statistical Analysis; Tape Recordings; *Test Construction; *Testing Programs; *Test Results
IDENTIFIERS	*Guidance Counselor Test

ABSTRACT

The two phases in the development and tryout of a Guidance Counselor Test to be added to the National Teacher Examinations Program are discussed. In Phase One, a 150-item written test and a 50-item written test based on taped stimulus material were produced. Each test consisted of five-choice multiple-choice questions. In Phase Two, the tests were administered during May-July 1972 to a sample of 129 college students who were completing work on a Master's Degree in Guidance and Counseling. Five institutions were represented in the study. In addition to taking the tests, each candidate completed a self-report Questionnaire Data Form. Several types of analyses were carried out, including two item analyses (one for the written test and one for the taped test). An intercorrelation matrix showing the correlations between the test and their various subscores was also obtained. Ten stepwise regression analyses were run, each being based on a continuous variable derived from the questionnaire and used as a dependent criterion variable, with the written and taped tests used as independent variables. Feedback from the candidates about the tests was, in general, positive. Results are provided as related to item and test characteristics, intercorrelations among the subscores on the written and taped tests, and relationships between questionnaire data and test scores. A table presents multiple correlations of combinations of taped and written subscores with criterion variables. The Questionnaire Data Form with scoring specifications and the Test Evaluation Form are provided. Tables in an appendix present intercorrelations among the taped test and written test, reliabilities of the subscores for the two tests, and zero-order correlations of the tests subscores with criterion variables. (DB)

ED 081782

TM 003 142

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TEST DEVELOPMENT REPORT

TDR73-1 APRIL 1973

DEVELOPMENT AND ANALYSIS OF A TAPED AND WRITTEN TEST FOR GUIDANCE COUNSELORS: A PILOT STUDY

Betty Humphry
Test Development Division, ETS

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Dorothy Urban

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Copyright © 1973 by Educational Testing Service. All rights reserved.

FILMED FROM BEST AVAILABLE COPY

TABLE OF CONTENTS

	Page
Acknowledgments	iii
Phase One: Development of the Test Materials	1
Development of Specifications for the Written Test	1
Item Writing for the Written Test and Related Validity Concerns	3
Development of Specifications for the Taped Test	4
Item Writing and Development of Taped Stimulus Materials	8
Phase Two: Tryout of the Test Materials, Analysis of Data, and Findings of the Study	12
Description of the Sample and Procedures for the Tryout and Analysis of Data	12
Feedback from Candidates About the Tests	15
Item and Test Characteristics	17
Intercorrelations Among the Subscores on the Written and Taped Tests	18
Relationships Between Questionnaire Data and Test Scores	20
Appendix A	28
Questionnaire Data Form	29
Scoring Specifications for the Questionnaire Data Form	31
Test Evaluation Form	33
Appendix B	36
Table B1: Intercorrelations Among the Taped Test and the Written Test	37
Table B2: Reliabilities of the Subscores for the Written and Taped Tests	37
Table B3: Zero-order Correlations of Taped and Written Test Subscores with Criterion Variables	38

ACKNOWLEDGMENTS

This project involved the efforts of many people both within and without Educational Testing Service (ETS). The many contributions of the Committee of Examiners in the production of the test specifications and in the subsequent development of the test materials should be acknowledged. Committee members include William Cash, University of Michigan; Robert Hoppock, New York University; John Krumboltz, Stanford University; Richard Rank, Georgia State University; and Lucille Stewart, Joliet Township (Illinois) High School. Dr. Charles McCracken, Newark State College, who made his logs of counseling situations available to Test Development, deserves special thanks, as do the colleges and students who cooperated with ETS in the tryout administrations of the tests. Appreciation is also extended to the drama class of Central High School of the Hopewell Valley Regional School District who participated in the final taping of the test and to the school administration.

Within the Test Development Division of ETS, the role of Aileen Cramer should particularly be acknowledged. Aileen worked with me on the development of the test materials and shared and helped to resolve many of the problems associated with the project. Test Development staff members who served as counselors and clients at varying stages of the taping include Robert Atwan, Harriet Bacon, Lawrence Beaber, Brigette Bell, Emily Bodenheimer, James Braswell, Gertrude Conlan, Aileen Cramer, Charles Daves, Jeanne Finelli, Harriett Frankel, Ralph Gallagher, Paulette Harris, Betty Humphry, Anna Jackson, Anna Mary Klein, Miriam Levin, Ann McAloon, Nona Johnson, Lora Olsen, George Pfeiffenberger, Mildred Rowen, Hachemi Saada, Gertrude Sanders, Anna

Schurter, Geraldine Smith, William Smith, Sally Stover, Steven Stupak, Diane Turner, Jefferson Wadkins, Michael Ward, Russell Webster, and Protase Woodford.

People from other Divisions who served as counselors or clients in the taping of the materials include William Harris, Arthur Kroll, Richard Majetic, William Parker and son Anthony, Keith Swisher, and Willie Turner. Andre Diaz recorded the directions for the test.

Other contributions to be acknowledged are those of Norman Wexler, who coordinated the statistical aspects of the study, provided the summary data of the regression analyses, and offered helpful suggestions in regard to the report; Richard Majetic and Andre Diaz, who arranged the tryout administration, procured the services of the Hopewell Valley drama class, and assisted with the final taping; John Rhubart, who coordinated operational procedures for the project; Ronald Kohl, who designed the system for processing and analyzing the data; Stewart Allen and the Production staff of the Test Development Division who processed the test materials; Arthur Kroll and Jean Reiss, who reviewed test materials; and Atty Van Hamel, who helped in the preparation of taped materials.

I am grateful, too, to Elizabeth Stewart, who reviewed the report in varying stages and made many helpful suggestions, including the format of the table summarizing the results of the regression analyses; and to Edward Masonis and Thomas Donlon, who also reviewed the report and provided valuable feedback.

PHASE ONE: DEVELOPMENT OF THE TEST MATERIALS

Development of Specifications for the Written Test

In April, 1971, a Committee of Examiners met at Educational Testing Service to develop specifications for a Guidance Counselor Test to be added to the offerings of the National Teacher Examinations (NTE) Program. The test was to be a two-hour test consistent with the other NTE Teaching Area/Field Examinations.

As an initial step, certain guidelines were set in regard to the population for which the test was intended. The test was to focus on guidance counselors rather than directors of guidance. It was to be geared to people who work in the public schools. It was to focus on the secondary school level rather than on the elementary school level, although a number of questions might have general applicability to both levels. It was assumed to be most appropriately taken at the end of the master's degree level of training.

Guidelines were also formulated for the framework to be used in the development of the test specifications. The specifications were to be centered on the competencies required of the counselor rather than on the particular content he or she had studied. These competencies, in turn, were to be related to desired student outcomes. In line with this, six major categories of counselor competencies and three major categories of student outcomes were agreed upon.

The six major areas of counselor competencies, together with representative subcategories, are:

I. Interpersonal Communication: Being able to, e.g., give accurate feedback of content, reflect feelings accurately, vary questions and statements to further communication, adapt communication skills to individual and cultural differences

II. Observing and Describing: Being able to, e.g., identify relevant behavior to be changed, infer feelings and intent from nonverbal cues, precisely describe crucial behaviors

III. Goal Setting: Being able to, e.g., help client state goals as tentative approximations, help client state goals as behavior changes likely to produce client satisfaction, help client state motives or values for proposed actions

IV. Systems for Changing Behavior: Being able to, e.g., prescribe actions that client or others can take to modify inappropriate behavior or elicit new behavior, such as arranging different reinforcement schedules or encouraging client base-rate taking; develop educational programs directed toward concerns such as anticipating future employment work habits or promoting harmonious interracial relations

V. Finding and Providing Information: Being able to, e.g., find information client cannot find, interpret information that might mislead client, help client avoid confusion by falsely considering projections as predictions

VI. Evaluating Outcomes: Being able to, e.g., judge success on basis of client attaining goals, design simple controlled experiments, use feedback to modify future procedures

The three major categories of student outcomes, together with representative subcategories, are:

1. Learn How to Identify and Anticipate Problems (e.g., group conflict resolution, causes of behavior or intent of action)
2. Learn How to Make Decisions Wisely (e.g., generate alternatives, find information)
3. Learn How to Take Constructive Action to Effect Change or Attain Personal Goals (e.g., overcome fears, make friends)

Item Writing for the Written Test and Related Validity Concerns

Prototype items for the written test were produced by the committee members at the April meeting. These were to be incorporated in a manual to be sent to item writers. Item writers would include both committee members and outside item writers the committee might recommend. The test was to be 150 items in length.

A part of the item-writing discussion centered on the use of modes of testing other than pencil and paper alone, and the suggestion was made that an audio portion be included in the test. Committee members felt that this mode would make it possible to obtain a different kind of information about the test candidate. He or she might, for example, listen to portions of an interview and be asked to identify feelings being expressed or to identify problems from what the client did or did not say.

Empirical validation was another concern raised in regard to the test. Pretesting of the test materials on groups of counselors was considered to be a desired procedure. This would also provide feedback about various item types that would aid in future test development should a combined written-audio test be produced.

Development of Specifications for the Taped Test

As a follow-up of the committee's recommendation that an audio portion of the test be considered, it was decided that a taped test would be developed on an experimental basis and tried out in conjunction with the two-hour, 150-item written test that was being developed. In line with this decision, two members of the Committee of Examiners met with Test Development personnel at ETS in September, 1971, to develop specifications for the taped materials.

A preliminary matter to be considered in the development of taped materials was the division of testing time between the regular written portion of the test and the taped portion, should a combined package eventually materialize in the NTE Program. For purposes of reliability, it was decided that the taped test ought to consist of at least 40-50 items and the written test of about 100 items. Subject to experimentation, division of the two hours of testing time for a final test that might be introduced into the Program was, therefore, to be as follows:

	No. Minutes	No. Items
Regular Written Portion	90	100
Taped Portion	30	40-50

The time limit and number of items for the experimental taped test was to be the same as that above, and the written test was to be a two-hour, 150-item test as previously indicated. The use of the 150-item written test in the tryout was undertaken for several reasons. First, the written test was originally conceived as an entity in itself and the test development process had proceeded on this basis. Secondly, relating the 150-item test as originally conceived by the committee to the experimental taped test would provide a better base for determining which categories of the written test specifications might most fruitfully be measured by written means and which

by taped means. Development of a 150-item written test would also, of course, permit introduction of the Guidance Counselor Test into the NTE Program without the taped portion if the latter appeared to measure nothing unique.

Basic ground rules set for the development of the audio portion of the test were:

1. The stimulus materials must be on tape.
2. The test items were to be of the five-choice, multiple-choice type built around the following:
   (a) a single statement by the client
   (b) a single statement by the counselor
   (c) a single client response and counselor response
   (d) multiple client responses and counselor responses (brief dialogue)
   (e) a longer excerpt that would lend itself to several questions
3. The test materials were to deal with face-to-face interpersonal situations.
4. The test questions were to be directed to (or refer back to) specific topics on the original (written test) specifications.
5. Sets of items that focused on a given stimulus must be limited to 4 or 5 per stimulus, and the stimulus excerpt must last no more than 1 1/2 minutes.

A further guideline in regard to the taped test materials was that appropriate balances be maintained among situation variables such as (1) counseling and consulting, (2) clients and counselors, (3) males and females, (4) blacks, whites, and Chicanos, (5) content and feelings, (6) individuals and groups,

and (7) settings to elicit descriptions of what is happening and settings to elicit judgments about what is happening.

In terms of behaviors to be tested, the items were to focus on behaviors that require the test taker to

1. identify counselor behaviors or techniques
2. evaluate appropriate counselor behaviors
3. identify client feelings
4. describe interaction patterns between counselor and client
5. select the client response the counselor is most likely to elicit
6. identify group process factors from an excerpt of group interaction

Two specific content categories that would provide a focus for stimulus materials were also delineated and are indicated below.

Adolescent Developmental Tasks	Emotional Triggers for Adults
1. Self-identity	1. Sex
2. Values	2. Motherhood
3. Sex role	3. Authority
4. Cultural group identification	4. Religion
5. Establishment of independence	5. Drugs
6. Peer relations	6. Criticism of professional competence

Some of the items in the test would be directed from these content categories to specific topics (counselor competencies) in the specifications for the written test. Others would be directed only to the counselor competencies in the written test.

A few examples may illustrate the use of the specifications. Suppose a given excerpt covered a group counseling situation composed of male and female students who were discussing the roles of men and women in society. One of the test questions might be directed toward what the counselor could do to further communication among the group. This item could be classified as follows:

Situation Variables	Counselor Competency Tested in Written Test Specifications	Behavior Tested in Taped Test Specifications	Content Focus in Taped Test Specifications
Counseling	To promote communication in groups	Evaluate appropriate counselor behaviors	Sex role
Males/Females			
Blacks/Whites			
Content			
Group			
Judgment			

Another excerpt might be based on an interaction between a male counselor and a female parent in which the parent is expressing hostility toward the counselor. One of the test questions might be directed toward the feelings the client seems to be expressing. This item could be classified as follows:

Situation Variables	Counselor Competency Tested in Written Test Specifications	Behavior Tested in Taped Test Specifications	Content Focus in Taped Test Specifications
Consulting	Precisely describe crucial behaviors	Identify client feelings	--
Female/Male			
White/White			
Feelings			
Individual			
Description			

Item Writing and Development of Taped Stimulus Materials

As a part of the meeting to develop specifications for the taped test, various item types were discussed. Two item types based on taped examples that one of the committee members had brought to the meeting were modified to adhere to the five-choice, multiple-choice format agreed upon. These were to serve as partial resource materials for test development. The same committee member also offered to make some tapes available to Test Development personnel to be used as possible sources of stimulus materials.

Several problems emerged in the development of the taped stimulus materials. One was trying to get existing taped materials of high enough technical quality that they could be reproduced. Related to this was the need to consider ethical concerns related to client-counselor confidences and to insure that no materials were used without permission of the parties involved. A further problem was that of obtaining sufficient variety in the types of counseling situations represented on the available tapes. The amount of time required to listen to the available taped materials was relatively high in relation to the number of excerpts that were isolated as potential stimulus materials for item writing.

Coupled with the above complexities were those involved in attempting to communicate the many facets of the test specifications to others who might write test items. It had originally been anticipated that some of the items would be written by outside item writers, but because of the problems that emerged in exploring the development of the taped materials, most of the items for the experimental test were written within ETS.

The difficulties associated with obtaining existing taped materials led to consideration of what other means might be used to arrive at realistic counseling situations that could be put on tape. A conference with several

people in the area who are concerned with the training of counselors was set up by the Test Development staff who were working on the test. Discussion at this meeting led to an offer by one of the persons in attendance, who had been associated with ETS through Project NOW, to make available his file of logs of actual counseling situations encountered by college students as part of their intern training as counselors. By going through these logs, Test Development personnel were able to compile a varied list of problem situations sufficient in number and degree of detail to write the necessary stimulus materials and related test questions.

A variety of types of stimulus materials and questions were written in line with the specifications for the taped test. These were grouped into five parts. The first four parts were based on brief taped excerpts and were designed to assess the test taker's ability to (1) identify client feelings, (2) identify counselor feelings, (3) select the probable reaction of a client to a particular counselor response, and (4) evaluate the appropriateness of given counselor responses. The fifth part of the test was based on longer excerpts, including group counseling situations. These excerpts were designed to measure a broader range of abilities including the ability to identify critical issues, theoretical constructs, counseling goals, and desired response patterns of counselors, as well as other abilities measured in the first four parts of the test.

Once the stimulus materials and questions for the test were prepared, the next step was to work out the mechanics of the testing situation. Decisions needed to be made about what parts of the test materials were to be put on tape, what form the directions were to take, what the balance of counselor/client roles was to be in terms of sex and ethnic background, what the timing

of the various item types was to be, and what the format of the test book should be. A script incorporating these decisions could then be prepared for use in putting the test materials together.

Preliminary tapes and test materials for review by the Committee of Examiners were prepared when the script was refined. The materials were also reviewed by others within ETS who have had training and experience in guidance and counseling. ETS staff members served as counselors and clients for the preliminary tapes. A high school drama class was to be used in the final taping, in accordance with recommendations by committee members who developed the specifications for the taped test.

Review of the preliminary test materials was positive, in general. The excerpts were considered to be characteristic of realistic counseling situations and to represent a balance of counseling situations. The item types in the taped test were considered to measure abilities not covered in the written test. The possible responses to items were considered to cover an array of plausible distracters.

Negative comments by reviewers related primarily to the more technical aspects of the test materials. For example, it was suggested that more time was needed for some of the questions and that minor modifications might be helpful in regard to the test directions. Other comments were made concerning the quality of the tapes in that it was not always possible to hear clearly all of the voices.

In view of the generally positive response to the taped test, a final script was prepared that took into account desired changes in the timing of some of the excerpts, the wording of some of the directions, and other suggestions that had been made by those who reviewed the test. Final taping of the experimental test was subsequently undertaken at a professional

recording studio. Representatives from the Test Development and Program Divisions of ETS served as counselors. The drama class of a local high school served primarily as the clients, with assistance from two other young people associated with ETS. The directions for the test were recorded by one of the NTE Program Directors.

PHASE TWO: TRYOUT OF THE TEST MATERIALS, ANALYSIS OF DATA, AND FINDINGS OF THE STUDY

Description of the Sample and Procedures for the Tryout and Analysis of Data

The tests were administered during May-July, 1972, to a sample of 129 college students (f = 80, m = 49) who were completing work on a Master's Degree in Guidance and Counseling. About one-third of the students had undergraduate majors in Education and about one-fifth had undergraduate majors in Psychology. The remainder had been primarily Social Science and Humanities majors.

Five institutions were represented in the study, including a New England college, two California universities, and two Southern universities. Administrative arrangements for the tryout of the materials were made through the assistance of members of the Committee of Examiners and other ETS contacts in the field. Score reports were to be provided to institutions and students who took part in the study.

The tests were administered under conditions similar to a regular administration in that a supervisor's manual outlining testing procedures had been prepared. Each candidate took the standard written test consisting of 150 five-choice, multiple-choice questions and a 50-item test (also five-choice, multiple-choice) based on taped stimulus material. Because of the amount of testing time required, however, the written test was taken in one sitting and the taped test in another sitting.

In addition to taking both tests, each candidate completed a self-report Questionnaire Data Form, which the Test Development staff had prepared, covering such items as (1) number of graduate hours in guidance and counsel-

ing, (2) expected level of training sought, (3) number of graduate hours devoted to intern/practicum experiences, (4) grade-point average in guidance and counseling courses, (5) grade-point average in intern/practicum courses, (6) years of teaching experience, (7) undergraduate major, (8) part-time counseling experiences as a teacher, (9) out-of-school counseling experiences, (10) experience directing extracurricular activities, (11) focus of counseling interest, (12) desired work setting, (13) practical/theoretical balance of training, (14) theoretical emphasis of counseling training, and (15) ethnic origin. A copy of the questionnaire may be found in Appendix A, page 29. Feedback from candidates was also requested through a Test Evaluation Form prepared by the Test Development staff. This form is also found in Appendix A, page 33.

Several types of analysis were to be carried out as a part of the study. First, two item analyses were to be requested--one for the written test and one for the taped test. These analyses would provide information concerning the discriminating power and the level of difficulty of the test questions and their concomitant statistical usefulness as prototype items for the operational tests.

An intercorrelation matrix showing the correlations between the tests and their various subscores was also to be obtained. The correlation data would be helpful in assessing the uniqueness of subsets of test questions and provide information useful for future test development. Subscore information was to be for internal use only and was to be based on the following breakdown of scores:*

*Numbers in parentheses indicate number of items per subscore.

Written Test	Taped Test
1. Interpersonal communication (32)	1. Identifying or describing counselor goals, responses, techniques (10)
2. Observing and describing (17)	2. Evaluating counselor behaviors, responses (15)
3. Goal setting (26)	3. Identifying client feelings, problems, critical issues (17)
4. Systems for behavior change (30)	4. Selecting client response counselor is most likely to elicit (8)
5. Finding and providing information (23)	
6. Evaluating outcomes (22)	

It should be noted that for purposes of analysis, the subscores on the taped test were compressed into four categories of behaviors because of the small numbers of items under some behaviors on the specifications for the taped test.

Regression analysis was a further means of exploring the data collected. Relating test score information to criterion variables covered by the questionnaire would provide information concerning criteria assumed to be relevant to performance as a guidance counselor and the related construct validity of the tests.

Ten stepwise regression analyses were run, each of which was based on a continuous variable that was derived from the questionnaire and used as a dependent criterion variable, with the written test (six subtests) and taped test (four subtests) used as independent variables. Sex was arbitrarily treated as a continuous (criterion) variable in the regression analyses. A verbal aptitude measure was used as a concomitant variable (covariate) in the regression analyses in order to control for institutional differences in ability.

Five categorical variables that were also derived from the questionnaire (Undergraduate Major, Race, Vocational Plans, Type of Training, and Desired Type of Counseling Setting) were studied separately by examining the distributions of the total tape and written scores for each of the five variables.

Feedback from Candidates About the Tests

Comments from candidates reported on the Test Evaluation Form (see Appendix A, page 33) are too numerous to report in detail. However, some highlights based on those who returned the form (N = 115) are reported below:

- About three-fourths of the candidates felt the taped test measured different kinds of abilities than those measured by the written test. Unique characteristics mentioned were "ability to respond to tone of voice, feeling," "more measurement of those factors crucial to counseling like empathy and comprehension," "listening...perhaps the most important skill in counseling," "ability to remember exactly what was said," "[covered] more practical aspects of counseling."
- Candidates were about evenly divided as to whether they felt they had sufficient time to answer the questions. Negative comments included "did not get sufficient time to analyze responses," "yes in first [parts], no in longer excerpts," "not enough time to read and consider the best answer," "on these type questions I really like to think in detail," "the time gave little time for evaluation of choice selections."

- Practically all of the candidates felt the directions were clear. In some cases, it was suggested that repeating the directions would be helpful.
- About three-fourths of the candidates felt the technical quality of the tape was satisfactory. Some comments were "good acting," "excellent right down to accents and intonation," "most of [the excerpts] were clearly heard, but a few were not clear," "sometimes I felt the people speaking spoke too quickly." Conditions of the testing situation were also mentioned--e.g., too hot, too many extraneous sounds, inappropriate sound control.
- The types of questions in the taped test that candidates appeared to consider most effective in measuring counselor ability were those related to the nature of or appropriateness of the counselor's response to a client.

In general, feedback from candidates was positive in nature, as is indicated above. Comments suggested that the taped test was essentially sound and useful for its intended purposes.

Two particular concerns in regard to the taped test might be noted. First of all, being certain that the listening conditions are good is an important consideration in administering the test, as indicated by candidates' responses. The timing of the various excerpts and questions is a more complex concern in that part of the measure of an effective counselor is his or her ability "to listen attentively," "to comprehend or perceive quickly," and the like. More data probably are needed as an information base in making decisions related to the timing of taped materials of this kind.

Item and Test Characteristics (N = 115)

	Number of Items	Raw Score Range	Raw Score Mean	Raw Score Standard Deviation	Mean Item Δ*	Mean Item r biserial	Reliability
Written Test	150	5-129	89.27**	25.46	11.04	.39	.94
Taped Test	50	8-45	31.65**	7.07	11.50	.43	.82

As may be noted from the mean raw score and mean Δ, the tests were somewhat easy for this population--the written test slightly more so than the taped test. If the tests had been of middle difficulty, the raw score means would have been 75 and 30 on the written and taped tests, respectively, and the mean delta for each test would have been 12.0. The raw score ranges and score standard deviations, relative to the numbers of items, are in accordance with those of most of the other Area Examinations in the NTE Program.

The mean r biserials indicate that the majority of items in the tests had the power to discriminate between candidates who scored high on the test and those who scored low on the test, using total score on each test as the criterion. The aim for tests related to education is for a mean lower limit r biserial of .30, and .40 is generally about the mean upper limit obtained.

Reliability coefficients were computed by a modification of Kuder-Richardson Formula 20. The estimated reliabilities for the two tests also reflect the tests' discriminating power, with .90 being the desired lower limit for a 150-item test related to education. Had the written test been 100 items, the number of items projected for the written portion of the

* Δ of 13.0 = 50% of population responding to the item answered it correctly. Δs lower than 13.0 indicate easier items--e.g., Δ12 = 60% answered item correctly.

** Scores on the written test reflect application of the standard penalty for guessing formula. Scores on the taped test are based on rights only.

operational final form, the estimated reliability would have been .91 (according to application of the Spearman-Brown formula). Since reliability is related to length of test, the reliability of .82 for the taped test of 50 items can be considered quite acceptable. Assuming that a final form test would be prepared that consisted of 100 items in the written portion and 50 items in the taped portion, the estimated reliability of the total test would be expected to meet the desired criterion of .90 or higher.

All of the data above should be considered only suggestive in the light of the limited sample of 115 cases. The lower-limit sample size for item analysis of tests of this kind is usually about 300 cases.

Intercorrelations Among the Subscores on the Written and Taped Tests

The correlation between scores on the total written test and scores on the total taped test was .81. This is a high correlation in relation to the estimated test reliabilities of .94 and .82, and it suggests that those who do well on the written test also do well on the taped test.

The intercorrelations among the subscores are less easily interpreted because of the variations in the number of items in each subscore and the small numbers of items upon which some subscores are based. The range is from 8 items for Subscore 4 of the taped test (see page 14) to 32 items for Subscore 1 of the written test. However, a few highlights are summarized below:*

- Correlations between Subscores 1, 2, 3, and 4 of the taped test and the total written test were .62, .73, .64, and .57,

*A complete table of intercorrelations, as well as the reliabilities of the various subscores, may be found in Appendix B, page 37.

respectively, which suggests that Subscore 2 (Evaluating counselor behaviors, responses) is more highly related to candidates' scores on the written test than any of the other subscores based on the tape. However, had the various subscores and the written test been perfectly reliable measures, the intercorrelations between Subscores 1, 2, 3, and 4 of the taped test and the total written test would have been .91, .96, .83, and .99, respectively, as revealed by correction of the intercorrelations for attenuation (or the effects of errors of measurement). With such correction, Subscore 4 (Selecting client response counselor is most likely to elicit) is correlated more highly with the total written test than is Subscore 2 (Evaluating counselor behaviors, responses).

- Subscores 1, 2, 3, 4, 5, and 6 of the written test correlated .74, .66, .74, .73, .63, and .65, respectively, with the total taped test, which suggests that Subscores 1 (Interpersonal communication), 3 (Goal setting), and 4 (Systems for behavior change) of the written test show the highest correlations with the total taped test. Correction of the intercorrelations for errors of measurement reveals that had the subscores on the written test and the total taped test been perfectly reliable, the intercorrelations between Subscores 1, 2, 3, 4, 5, and 6 of the written test and the total taped test would have been .93, .86, .92, .89, .85, and .88, respectively. Written Test Subscores 1 (Interpersonal communication)

and 3 (Goal setting) thus continue to show the highest correlation with the total taped test. However, Subscore 6 (Evaluating outcomes) is almost as highly correlated with the total taped test as is Subscore 4 (Systems for behavior change).

Relationships Between Questionnaire Data and Test Scores

Study of the five categorical variables derived from the Questionnaire Data Form revealed that two of the variables, Undergraduate Major and Race, produced significant differences in terms of mean test score performance. There was a tendency for candidates with undergraduate majors in Humanities, Psychology, and the Social Sciences to do somewhat better on both tests than candidates with undergraduate majors in Education, Science-Mathematics, or Languages. The 100 white candidates who took the tests tended to score higher on both tests than did the 15 black candidates.

Regression analyses based on 105 cases from four institutions (incomplete information was available for the fifth institution) revealed that parts of the taped test in conjunction with parts of the written test are more related to the criterion variables than either the taped test or the written test alone. (The dependent criterion variables were not necessarily experimentally independent of each other.) Appendix B, page 38, shows the intercorrelations among the covariate, the predictors, and the criterion variables prior to the regression analyses.

The table below lists the ten dependent criterion variables and shows the multiple correlations observed using the best combination of written and tape subscores with a verbal aptitude measure partialed out. Criteria to be met for continuing or stopping the stepwise regression analyses, based

on proportion of the variance accounted for by the measures, were .0125 or more to enter a measure and .00975 or less to drop a measure. The scoring scheme for each of the variables in the questionnaire may be found in Appendix A, page 31.

Table 1 -- Multiple Correlations of Combinations of Taped and Written Subscores with Criterion Variables

Dependent Criterion Variable	Best Single Predictor of Criterion, Together with Added Predictors WS = Written Subscore; TS = Taped Subscore	Multiple Correlations (with verbal measure partialed out)
Graduate Hours in Guidance and Counseling	WS 5 (Finding and providing information)	.27
	TS 2 (Evaluating counselor behaviors, responses)	.31
	WS 2 (Observing and describing)	.35**
Expected Level of Training Sought	WS 6 (Evaluating outcomes)	.18
Graduate Hours in Intern/Practicum Courses	WS 2 (Observing and describing)	.19
	TS 4 (Selecting client response counselor is most likely to elicit)	.25
	TS 1 (Identifying or describing counselor goals, responses, techniques)	.28*
Grade-Point Average Graduate Courses in Guidance/Counseling	TS 2 (Evaluating counselor behaviors, responses)	.17
	WS 2 (Observing and describing)	.23
	WS 5 (Finding and providing information)	.27
	TS 4 (Selecting client response counselor is most likely to elicit)	.30
Grade-Point Average in Intern/ Practicum Courses	TS 3 (Identifying client feelings, problems, critical issues)	.24
	WS 2 (Observing and describing)	.27
	WS 5 (Finding and providing information)	.31
	WS 3 (Goal setting)	.33
	TS 4 (Selecting client response counselor is most likely to elicit)	.37*
Years of Teaching Experience	TS 3 (Identifying client feelings, problems, critical issues)	.34**
Part-time Counseling Experiences as a Teacher	WS 1 (Interpersonal communication)	.16
	WS 4 (Systems for behavior change)	.26
	WS 5 (Finding and providing information)	.30
	TS 4 (Selecting client response counselor is most likely to elicit)	.33*
Direction of School Extracurricular Activities/Counseling with Youth Groups in Community	WS 5 (Finding and providing information)	.20
	TS 1 (Identifying or describing counselor goals, responses, techniques)	.31
	WS 4 (Systems for behavior change)	.35**
Theoretical versus Practical Training Emphasis	WS 1 (Interpersonal communication)	.31
	WS 6 (Evaluating outcomes)	.34**
Sex	TS 3 (Identifying client feelings, problems, critical issues)	.15
	WS 1 (Interpersonal communication)	.29
	WS 2 (Observing and describing)	.36**

*Significant at .01 level

**Significant at .05 level

A number of observations of possible interest may be made on the basis of the table. Among these are:

- Partialing out a verbal measure (or trying to eliminate the effect of verbal ability upon test performance) revealed that the taped test makes more unique contributions to the total Guidance Counselor Test than is suggested by the relatively high intercorrelation between the written test and the taped test indicated on page 18. The taped test contributes to the prediction of 8 of the 10 criterion variables listed.
- Performance on Subscore 3 of the taped test (Identifying client feelings, problems, critical issues) was the best single predictor of Grade Point Average (GPA) in Intern/Practicum courses, assuming the self-report information about GPA is accurate. Higher grades tended to be associated with higher scores on Subscore 3.
- Performance on Subscore 3 of the taped test (Identifying client feelings, problems, critical issues) was also significantly related to Years of Teaching Experience. There was a tendency for those who achieved higher scores on Subscore 3 to have fewer years of teaching experience. However, since about 40 per cent of the sample had no teaching experience, it is difficult to assess the precise nature of the relationship between the two measures. None of the subscores on the written test had significant predictive value in regard to Years of Teaching Experience.

- There was a tendency for women to achieve higher scores than men on Subscore 3 of the taped test (Identifying client feelings, problems, critical issues). Subscores 1 and 2 of the written test (Interpersonal communication and Observing and describing) also had predictive value in regard to sex of the test taker.
- Subscore 1 of the written test (Interpersonal communication) is the best single predictor of Theoretical versus Practical Emphasis in counselor training. Those who scored high on Subscore 1 of the written test tended to indicate that their training had been more theoretical than practical. Subscore 6 of the written test (Evaluating outcomes) also contributes to the prediction of Theoretical versus Practical Training Emphasis.
- The best single predictor of Part-time Counseling Experiences as a Teacher was Subscore 1 of the written test (Interpersonal communication). Those who had had such experience tended to receive lower scores on this subscore. Subscores 4 and 5 of the written test (Systems for behavior change and Finding and providing information) and Subscore 4 of the taped test (Selecting client response counselor is most likely to elicit) also contribute to the prediction of this variable. The specific nature of the relationships between these measures is difficult to assess because, as previously indicated, such a sizable proportion of the sample had had no teaching experience.

- Those who indicated that they had directed school extra-curricular activities or had had experience working with youth groups in the community tended to receive higher scores on Subscore 5 of the written test (Finding and providing information). Subscore 1 of the taped test (Identifying or describing counselor goals, responses, techniques) and Subscore 4 of the written test (Systems for behavior change) also contribute to the prediction of these variables. Again, however, precise interpretation of the data is compounded by the need to take into account the sizable proportion of the sample which had had no teaching experience and the fact that two variables were combined in the regression analysis.
- Subscore 2 of the written test (Observing and describing) was the best single predictor of number of graduate hours in intern/practicum courses. Those who received higher scores on this subscore tended to have more graduate hours in such courses. Taped Subscores 1 and 4 (Identifying or describing counselor goals, responses, techniques and Selecting client response counselor is most likely to elicit) also had predictive value in regard to this variable.
- Subscore 5 of the written test (Finding and providing information) was the best single predictor of number of graduate hours in guidance and counseling. Higher scores were associated with higher number of graduate

hours. Subscore 2 of the written test (Observing and describing) and Subscore 2 of the taped test (Evaluating counselor behaviors, responses) also contribute to the prediction of number of graduate hours in guidance and counseling.

In interpreting the data above, one should keep in mind the fact that intercorrelations among the dependent criterion variables are ignored. It is quite likely that many of them, if not all, are intercorrelated and that measurement of the same thing is being repeated from variable to variable. The fact that the questionnaire data are based on self-report information is another limitation.

Despite the number of limitations associated with a pilot study of the sort undertaken in this investigation, there appears to be evidence that the taped portion of the Guidance Counselor Test provides information that is unique to some of that provided by the written portion of the test and that this information is associated with important types of educational measures.

APPENDIX A

_____
ID
(from card given to you)

QUESTIONNAIRE DATA FORM

(Guidance Counselor Tests)

The information below is being collected for research purposes. It will not affect your score. Your cooperation in completing the questionnaire will be very much appreciated.

Name _____
Last First Middle Init.

Date of Birth _____ Sex _____
Month Date Year

Home Address _____

_____

Institution _____

DIRECTIONS: Please answer the following questions by circling the letter of the response that best applies to you. Mark only one response for each question.

1. a) If your credits are in semester hours, how many semester hours of graduate work in guidance and counseling will you have completed by June, 1972?

(A) 10 or less (B) 11-15 (C) 16-20 (D) 21-25 (E) 26-30 (F) 31+

b) If your credits are in quarter hours, how many quarter hours of graduate work in guidance and counseling will you have completed by June, 1972?

(A) 10 or less (B) 11-15 (C) 16-20 (D) 21-25 (E) 26-30 (F) 31+

2. At which of the following points do you expect to terminate your training in guidance and counseling?

(A) Masters Degree (B) Masters Degree + 30 hours (C) Doctorate

3. a) If your credits are in semester hours, how many semester hours of your graduate work will have been devoted to intern/practicum experience by June, 1972?

(A) 3 or less (B) 4-7 (C) 8-12 (D) 13+

b) If your credits are in quarter hours, how many quarter hours of your graduate work will have been devoted to intern/practicum experience by June, 1972?

(A) 3 or less (B) 4-7 (C) 8-12 (D) 13+

4. Using the following scale, indicate your grade-point average in your graduate courses in guidance and counseling to date. (A = 5; B = 4; C = 3)

(A) 5.0 (B) 4.7-4.9 (C) 4.4-4.6 (D) 4.1-4.3 (E) 3.8-4.0 (F) 3.7 or below

5. Using the same scale, indicate your grade-point average in your intern/practicum courses.

   (A) 5.0 (B) 4.7-4.9 (C) 4.4-4.6 (D) 4.1-4.3 (E) 3.8-4.0 (F) 3.7 or below

6. How many years of teaching experience have you had?

   (A) None (B) 1 (C) 2 (D) 3 (E) 4 (F) 5 (G) 6 (H) 7+

7. In which of the following categories would you classify your undergraduate major?

   (A) Education (B) Psychology (C) Science-Mathematics (D) Social Studies
   (E) Humanities (F) Languages (G) Other

8. Did your work as a teacher, if any, include any part-time counseling of students?

   (A) Yes (B) No

9. Have you had any counseling experiences other than those required as part of your graduate training or as a teacher (e.g., working with youth groups in the community)?

   (A) Yes (B) No

10. Did your work as a teacher, if any, include directing any extracurricular activities?

    (A) Yes (B) No

11. In which of the following areas of counseling are you most interested?

    (A) Educational planning (B) Vocational planning
    (C) Dealing with personal problems

12. In which of the following settings do you plan to work?

    (A) Elementary school (B) Secondary school (C) Junior or community college
    (D) Four-year college (E) Business-industry (F) Private agency
    (G) Community agency, rehabilitation, employment services, YMCA, YWCA, etc.

13. Which of the following best describes the practical/theoretical balance of your training?

    (A) More theoretical than practical (B) More practical than theoretical
    (C) Fairly evenly balanced between theory and practice

14. How would you describe the primary theoretical emphasis of your training?

    (A) Rogerian (B) Behavioral (C) Directive (D) Psychoanalytic
    (E) Rational (F) I consider my training to be eclectic.

15. Which of the following best describes your ethnic origin?

    (A) American Indian (B) Black, Afro-American (C) Mexican-American
    (D) Oriental (E) Puerto Rican (F) Spanish-American
    (G) White (H) Other

Scoring Specifications for the Questionnaire Data Form

Item #	A	B	C	D	E	F	G	H	Omit
1 (a)	12	20	27	34	42	50			25
1 (b)	08	13	18	23	28	33			
2	1	2	3						2
3 (a)	02	08	15	22					9
3 (b)	02	06	10	15					
4	5	4	3	2	1	0			2
5	5	4	3	2	1	0			2
6	0	1	2	3	4	5	6	8	2
7*	1	2	3	4	5	6	7		0
8	2	4							3
9**	2	4							3
10**	2	4							3
11*	1	2	3						0
12*	1	2	3	4	5	6	7		0
13	3	1	2						2
14*	1	2	3	4	5	6			0
15*	1	2	3	4	5	6	7	8	0

*Categorical variables

**A single score was obtained for questions 9 and 10 in combination. It was computed by adding the weights assigned to the options that the subject selected and subtracting the sum from 9.

_____
ID No. (Optional)

TEST EVALUATION FORM

(Guidance Counselor Tests)

To the Candidate: It will be of help in planning future test development activity if you will record on this form your reactions to the tests you have taken. Your assistance is very much appreciated.

1. Did you feel that the taped part of the test measured different kinds of abilities than those measured by the written test? ____Yes ____No

   Comments: _____

   _____

   _____

   _____

2. Was there sufficient time for you to answer the questions in the taped test? ____Yes ____No

   Comments: _____

   _____

   _____

   _____

3. Were the directions for the taped test clear? ____Yes ____No

   Comments: _____

   _____

   _____

   _____

4. Was the technical quality of the tape such that the voices were clearly heard? ____Yes ____No

   Comments: _____

   _____

   _____

   _____

(Over)

5. What kinds of questions in the taped test did you think were most effective in measuring counselor ability?

_____

_____

_____

_____

6. Was there sufficient time for you to answer the questions on the written (longer) test? ____Yes ____No

Comments: _____

_____

_____

_____

7. What were your reactions to the written (longer) test as a whole?

_____

_____

_____

_____

Additional Comments About Either Test:

APPENDIX B

Table B1

Intercorrelations Among the Taped Test and the Written Test

	Total Tape	TS 1	TS 2	TS 3	TS 4	Total Written	WS 1	WS 2	WS 3	WS 4	WS 5	WS 6
TS1	.75											
TS2	.85	.50										
TS3	.86	.54	.59									
TS4	.69	.44	.47	.47								
Total Written	.81	.62	.73	.64	.57							
WS1	.74	.54	.70	.59	.50	.88						
WS2	.66	.48	.59	.51	.50	.81	.70					
WS3	.74	.58	.64	.59	.52	.90	.79	.67				
WS4	.73	.54	.67	.57	.52	.91	.73	.72	.78			
WS5	.63	.50	.58	.48	.44	.84	.67	.60	.70	.71		
WS6	.65	.53	.55	.54	.43	.79	.57	.55	.66	.71	.66	

Table B2

Reliabilities of the Subscores for the Written and Taped Tests
(N = Number of Items)

TS1 N=10	TS2 N=15	TS3 N=17	TS4 N=8	WS1 N=32	WS2 N=17	WS3 N=26	WS4 N=30	WS5 N=23	WS6 N=22
.49	.61	.63	.35	.78	.70	.77	.83	.73	.67

Table B3

Zero-order Correlations of Taped and Written Test Subscores with Criterion Variables

	VERB APT INST MEAN	T. SUB. 1	T. SUB. 2	T. SUB. 3	T. SUB. 4	W. SUB. 1	W. SUB. 2	W. SUB. 3	W. SUB. 4	W. SUB. 5	W. SUB. 6	GRAD. HR. COUNS.	EXPECT. LEV. TRG.	INTERN HR.	GPA COUNS. CRS.	GPA INTERN CRS.	YR. TEACH. EXP.	P-T. COUNS. EXP.	DIR. EXTRA-CURR.	THEOR./ PRACT. TR.
T. Subscore 1	50																			
T. Subscore 2	39	50																		
T. Subscore 3	40	56	60																	
T. Subscore 4	28	43	47	47																
W. Subscore 1	33	57	70	63	49															
W. Subscore 2	42	52	60	52	49	74														
W. Subscore 3	40	61	66	61	54	81	70													
W. Subscore 4	58	53	63	56	48	72	70	75												
W. Subscore 5	45	48	57	50	45	70	66	72	70											
W. Subscore 6	51	56	52	57	43	63	65	70	69	69										
Grad. Hr. Couns.	40	24	38	32	27	29	23	38	38	40	32									
Expect. Lev. Trg.	07	09	09	07	09	16	08	18	10	19	19	15								
Intern Hr.	80	32	30	30	27	20	24	25	46	34	39	42	13							
GPA Couns. Crs.	-04	04	14	13	13	02	06	06	04	11	05	18	09	01						
GPA Intern Crs.	33	30	24	34	27	18	14	17	26	29	22	44	12	50	29					
Yr. Teach. Exp.	-35	-33	-35	-43	-27	-30	-26	-27	-29	-21	-31	-15	-02	-31	-22	-39				
P-T. Couns. Exp.	05	16	09	14	15	16	15	07	-00	-02	07	-02	-21	-05	02	05	-21			
Dir. Extracurr.	19	-04	07	08	01	10	09	08	08	26	16	13	24	14	01	01	06	-24		
Theor./Pract. Tr.	-35	-01	08	02	09	16	04	06	-05	-06	-13	-16	-05	-27	-12	-02	-06	-03	05	
Sex	07	02	00	16	10	-09	09	-06	04	-07	07	-03	-28	-05	11	04	-12	14	-20	-00